Combination of differential expression and co-expression network analyses identify novel conserved age-associated changes among different tissues

Caio M.P.F. Batalha[1], André Fujita[2], Nadja C. de Souza-Pinto[1]*

[1]Depto. de Bioquímica, Instituto de Química, and [2]Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, SP, Brazil, 05508-000

*To whom correspondence should be addressed at nadja@iq.usp.br

Running title: Tissue-specific gene expression changes with age

Keywords: aging, gene expression, co-expression network analysis

Abstract

Changes in gene expression patterns with age are observed in all organisms and tissues studied thus far. However, very little overlap is observed when comparing absolute gene expression changes across tissues, which could be interpreted as a lack of commonly affected pathways in aging. Considering that a given gene function can be altered by changes in its absolute expression level but also by changes in co-expression networks, we analyzed differences with age in absolute gene expression levels as well as using several network analysis techniques, which quantify co-expression levels in weighted networks. This revealed that some tissues display great changes in gene co-expression, but not in absolute gene expression, and vice-versa. By combining these altered genes with age in different metrics and network clustering, we show that most of the processes modulated by age are similar across tissues, with the most prominent being modules enriched with terms related to mitochondria, immune system, autophagy, protein catabolism and folding, RNA splicing and ribosomal biogenesis, with all the terms in this last group falling under the same connectivity modules. Further analyses of module preservation with age reinforced the conservation of affected processes across tissues under the light of gene co-expression. Interestingly, despite the similarities in terms of enriched processes, genes changing across tissues in any metric showed remarkable lack of overlap, indicating that although some processes seem to display clear conserved changes across tissues, the actual genes associated with these processes that are altered seem to be tissue-specific, highlighting the importance of interrogating processes instead of individual genes in systems level comparisons. Our results show that when combining analyses of absolute gene expression changes and gene co-expression changes some clear age-associated patterns emerge across tissues. Notably, the prominence of modules with terms that can be mechanistically linked to mTOR indicates a possible interpretation of how mTOR inhibition mediates lifespan extension.

Introduction

Aging is a multifactorial process that leads to alterations in several essential pathways resulting in increased risk of death. With some notable exceptions, where the detection of age-related decline has proven elusive[1], aging affects most known organisms, including unicellular ones. But despite its broad nature, the underlying mechanisms leading to the aging phenotypes are not well established. Nonetheless, much progress has been made in the last decades in understanding the cellular changes associated with aging and in identifying some well-conserved hallmarks[2], which provide a helpful framework for organizing the various age-associated alterations, even though they do not shed much light on the mechanisms behind the aging process[3].

The development of high-throughput technologies has allowed the generation of large amounts of data for gene expression, epigenetic markers, protein levels, and other biological endpoints. The possibility of gathering data from large numbers of individuals per age group, or longitudinally, provides unique opportunities for researchers to explore system-level patterns and trajectories of biological endpoints during the aging process, enabling the search for conserved patterns in these biological endpoints among organisms and tissues that help to explain the aging process. For instance, analysis of methylation status of a small amount of CpG islands, in what is known as the epigenetic clock, allows predicting an organism's age with remarkable accuracy, in what has been a crucial advancement in the search for comprehensive biomarkers of aging[4-8]. Notably, some of these epigenetic clocks are pan-tissue, developed with data generated from several tissues and representing a pattern of methylation changes that is common across several tissues[4; 6]. The same approach was used to generate a clock capable of predicting the maximum lifespan of different species, highlighting the essential role of DNA methylation changes in aging across evolution[9].

Despite the success in finding conserved DNA methylation patterns across tissues and organisms, finding such patterns with gene expression data has proven to be more challenging. Several studies have compared age-related changes in gene expression across tissues in humans[10-14] and rodents[10; 15], with inconclusive results. While some report no overlap in differentially expressed genes with age between tissues, others find some subtle commonalities between a few tissues. Therefore, the question of whether tissues age differently or if there are common aging alterations reflected in conserved tissue-independent transcriptomic changes remains.

Gene expression profiles can vary not only in their absolute expression values, but also in the genes' co-expression patterns. Gene co-expression is typically represented as a correlation that measures how coordinated the expression of two genes are. Gene expression levels and gene co-expression patterns, however, do not have a direct relationship, and genes that vary in one measurement may not vary in the other. Given that previous works on gene expression and aging focused mostly on analyzing absolute gene expression changes, it is relevant to ask whether genes that do not change expression levels with age may have altered co-expression patterns. In principle, a gene does not need to have a change in its expression to result in altered activity. Systemic changes in signaling and overall organellar and cellular function can result from altered co-expression pattern of genes under specific conditions, even if their absolute expression values are not altered. Therefore, we decided to analyze gene expression and gene co-expression separately and contrast the two measurements with the aim of exploring how they behave during aging. To carry out these analyses, we used correlation networks.

Weighted co-expression networks, constructed with pairwise gene expression correlations, provide a robust framework for analyzing system-level changes between different conditions[16]. In these networks, genes are represented as nodes, and vertices indicate pairwise relationships between genes. These relationships – essentially, gene expression correlations – provide a weighted measurement of the strength of the interaction between two genes. Using weighted networks to represent gene co-expression data is preferable to using discrete networks, that rely on hard thresholds, since they preserve the continuous nature of the information. Gene co-expression networks have been successfully used in analyzing biological data, including in the aging context[17-20]. Moreover, networks allow the calculation of different network structural measurements that can, then, serve as metrics for comparing different networks. One of the most basic types of network topology measurement is network connectivity (or degree centrality), which represents the sum of the network adjacencies (vertices) for each gene. Therefore, degree centrality in a gene co-expression network represents a measurement of each gene's general correlation with all other genes in the network. Another helpful metric is eigenvector centrality, which scores nodes based on the centrality scores of the other nodes they are connected to, such that a node with a high eigenvector centrality will be connected with nodes with high scores. Together, these network parameters reflect distinct biological information than absolute changes in gene expression levels.

In this work we analyzed changes in gene expression and co-expression networks during aging using RNA-Seq data from eight tissues from the Genotype-Tissue Expression (GTEx) project[21]. Our results demonstrate that changes in gene co-expression, measured by several co-expression network analyses, reveal alterations that cannot be revealed by analyzing changes in absolute gene expression alone, and that when both gene expression and gene co-expression data are taken together, a more robust understanding of the regulation of gene expression during aging arises, showing conserved processes being modulated with age across different tissues in humans.

Results

Analyses of tissue-specific differentially expressed and connected genes: For this study we used TPM-normalized RNA-Seq data from the GTEx project (version 8), which has whole transcriptome data for up to 54 tissues from 948 *postmortem* donors. To maximize statistical power, we aggregated samples by tissue type (the SMTS variable in the GTEx sample attributes) and selected only the tissues with at least 800 samples. This resulted in eight tissues being selected: blood (929 samples), brain (2642 samples), adipose tissue (1204 samples), muscle (803 samples), blood vessel (1335 samples), heart (861 samples), skin (1809 samples), and esophagus (1445 samples). The data was processed as described in the methods section.

For the gene expression analysis, we fit a linear mixed model using the nlme R package[22; 23] to see the effect of age on TPM levels, using as covariates the variables sex, race, death type (Hardy scale), ischemia time, and experimental batch. For the differential connectivity analysis, before calculating centrality measures, we corrected the dataset for the same confounding factors used in the previous linear mixed model using ComBat[24]. Figure 1A show as general schematics of the data processing.

The batch-corrected dataset was divided into 6-year age ranges, except for the last one of only 3-years (Figure 1B). By dividing the dataset into intervals of six years, we guarantee that the smallest number of samples among all the age ranges and all tissues is above the minimum

recommended number of samples for WGCNA[25] network construction (20 samples). In other words, the smallest age bracket at Figure 1B needs to have at least 20 samples. With this distribution, the smallest dataset is for heart, for the age range 26 - 31 years old, with 21 samples. The last age range was set to 3-years because no division of the whole dataset by a single number could maintain the previously mentioned criterion without using a considerable number of years as an age range. Using age ranges larger than 6-years would decrease the number of points as to make the subsequent regression analysis unviable. Nonetheless, since the GTEx dataset has an unbalanced number of samples per age, with relatively fewer donors at younger ages and many older ones, the smaller final age range, of 3-years, still has 79 samples in the tissue with less samples in this age range, muscle, providing enough samples for network construction (Fig. 1B). Finally, the gender distribution among the tissues is similar (Figure 1C), with around 70% of the samples coming from males.

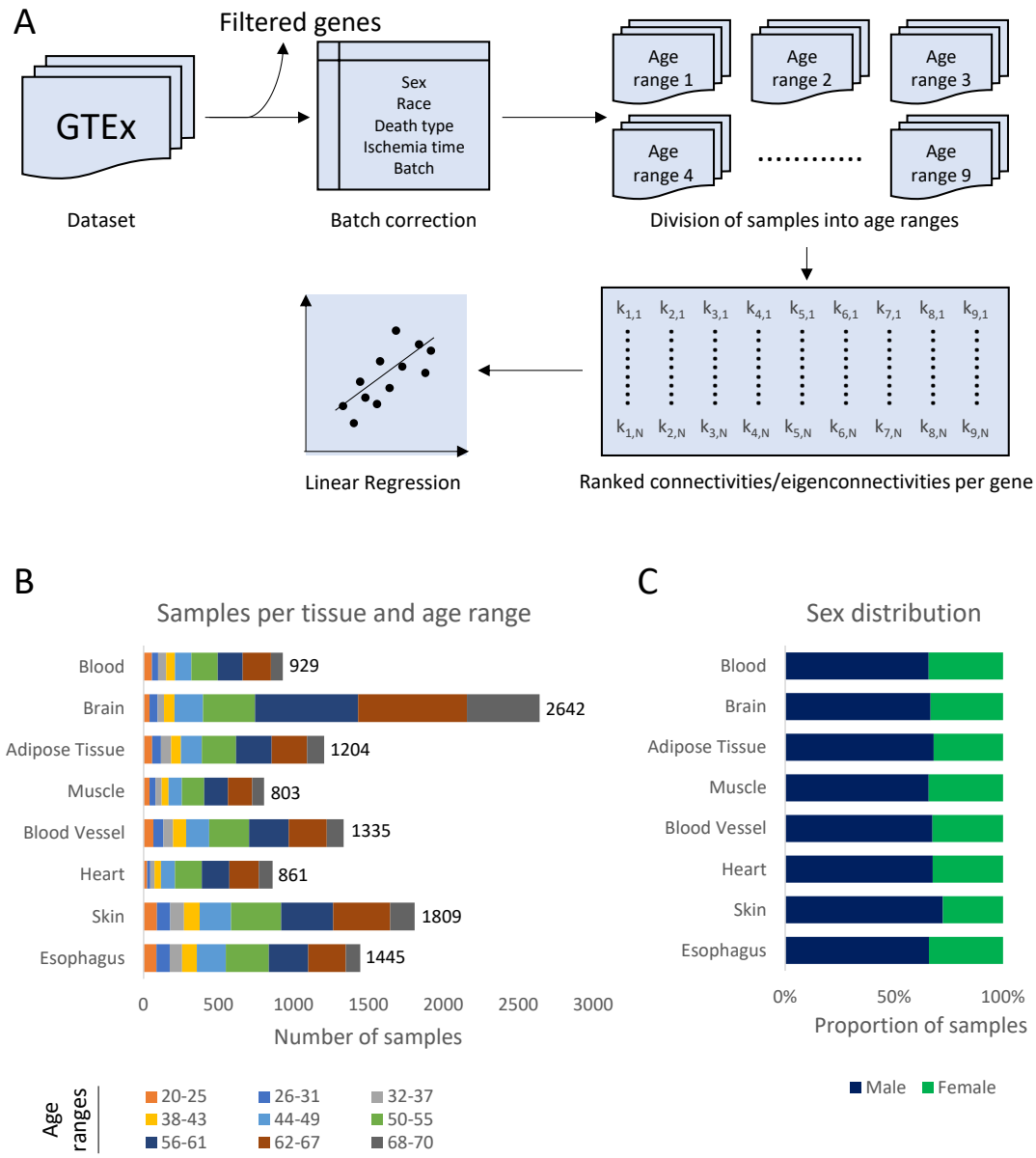


Figure 1 – Experimental design and dataset description. A) Schematics of the dataset partitioning into age ranges and processing for calculation of differential connectivity and eigenconnectivity. B) Number of

samples per tissue per age range, comprised of 6-year intervals, except for the last group, of 3 years. The total number of samples per tissue is shown to the right. C) Sex distribution of samples per tissue.

The datasets from the eight tissues were used to calculate the soft-threshold power for the weighted adjacency matrix using WGCNA. Since every tissue was divided into nine age ranges, the soft-threshold power was estimated for each of these nine networks in each tissue. Then, we selected a single power per tissue that could satisfy all nine networks simultaneously and be further used for consensus network construction. The powers selected for blood, brain, adipose tissue, muscle, blood vessel, heart, skin, and esophagus were 6, 6, 9, 12, 8, 12, 6, and 6, respectively. We then built adjacency matrices for all nine networks in the eight tissues using these powers with WGCNA. Network connectivity (node strength) and eigenvector centrality were calculated from the adjacency matrices using the igraph package[26]. This resulted in nine connectivity measures per gene in each tissue across the nine age ranges. A direct comparison of connectivity measures in different networks is challenging since different sample size, and technical variations can lead to different numerical values that do not necessarily reflect a difference in the underlying biology. Thus, we opted to compare the ranks of each connectivity. Therefore, we ranked the connectivities inside each network and applied a linear regression on the nine age ranges to detect significant differences in ranked connectivities for each gene.

Figure 2A shows the results of the differential expression, differential connectivity, and differential eigenvector centrality analyses. Genes were considered altered in any of these metrics only if the regression analyses returned an adjusted p-value (FDR[27]) < 0.05. The full regression results for the three metrics are presented in Supplemental Data 1. The amount of differentially expressed genes (DEGs), differentially connected genes (DCGs) and differentially eigen connected genes (DECGs) varied considerably by tissue. Altogether, there seems to be a tendency towards up-regulation in DEGs and DCGs, with a more pronounced tendency in DEGs and a more discrete tendency in DCGs, while DECGs display mixed results.

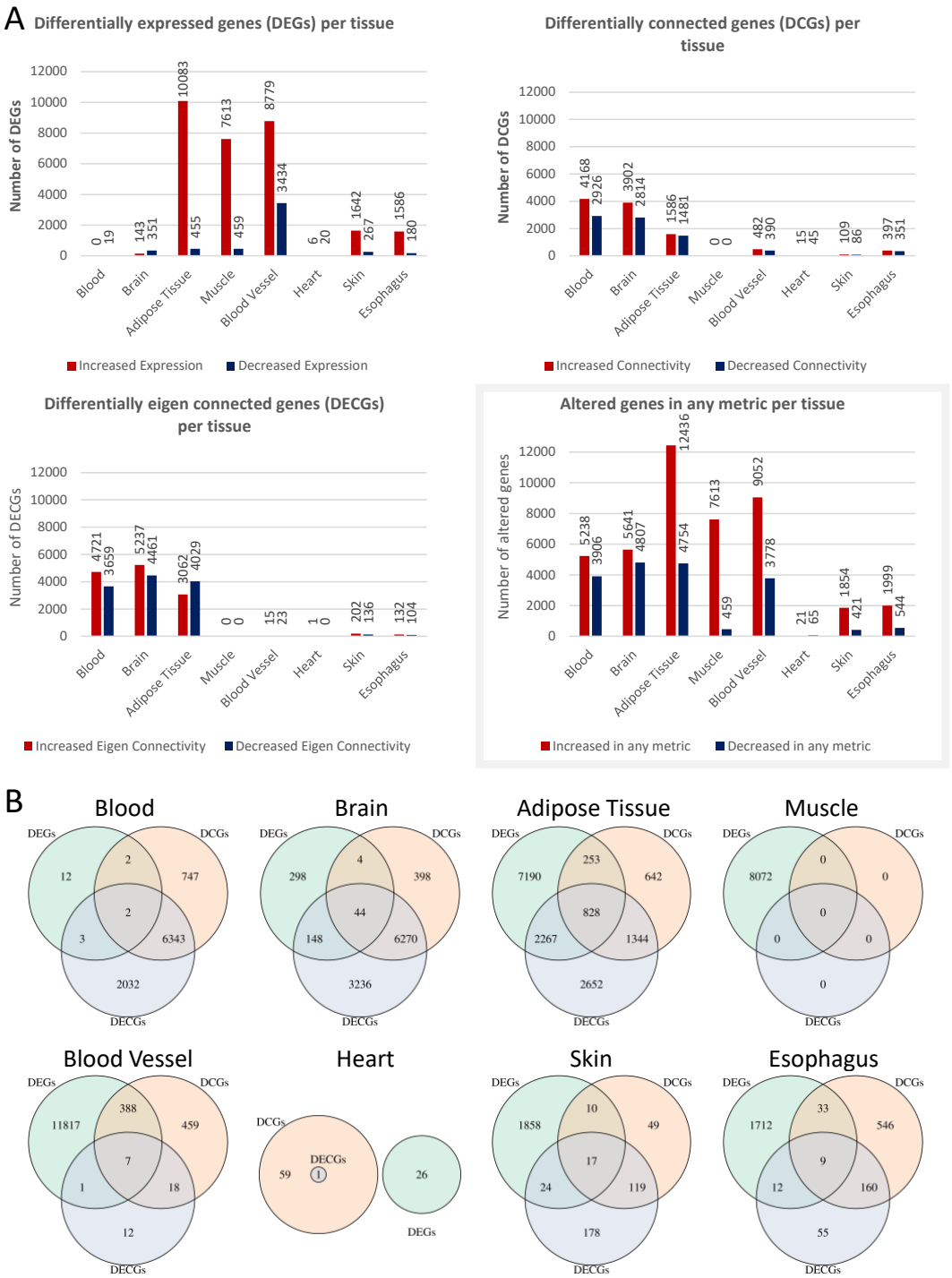


Figure 2 - Differentially expressed genes (DEGs), differentially connected genes (DCGs), and differentially eigenconnected genes (DECGs) with age per tissue. A) DEGs, DCGs, and DECGs per tissue detected by linear regression. The red and blue bars represent genes with positive and negative slopes, respectively. The last panel shows the number of genes with altered parameters in at least one of the three metrics. All metrics considered FDR < 0.05. B) Venn diagrams of intersections between DEGs, DCGs, and DECGs per tissue.

All tissues had a considerable number of genes changing with age in at least one metric, except the heart, which demonstrated relative stability in all three (Figure 2A, lower right panel). Interestingly, most tissues demonstrated relative stability in one type of analysis and great variability in others. Blood, brain, muscle, blood vessel, skin, and esophagus all fit into this category, having many altered genes in either gene expression or the centrality analyses, but not in both simultaneously. On the other hand, DCGs and DECGs showed some concordance in most cases, suggesting that, indeed, DEGs and DCGs/DECGs are measuring two different phenomena. Adipose tissue was the only tissue that simultaneously showed great changes in all three metrics.

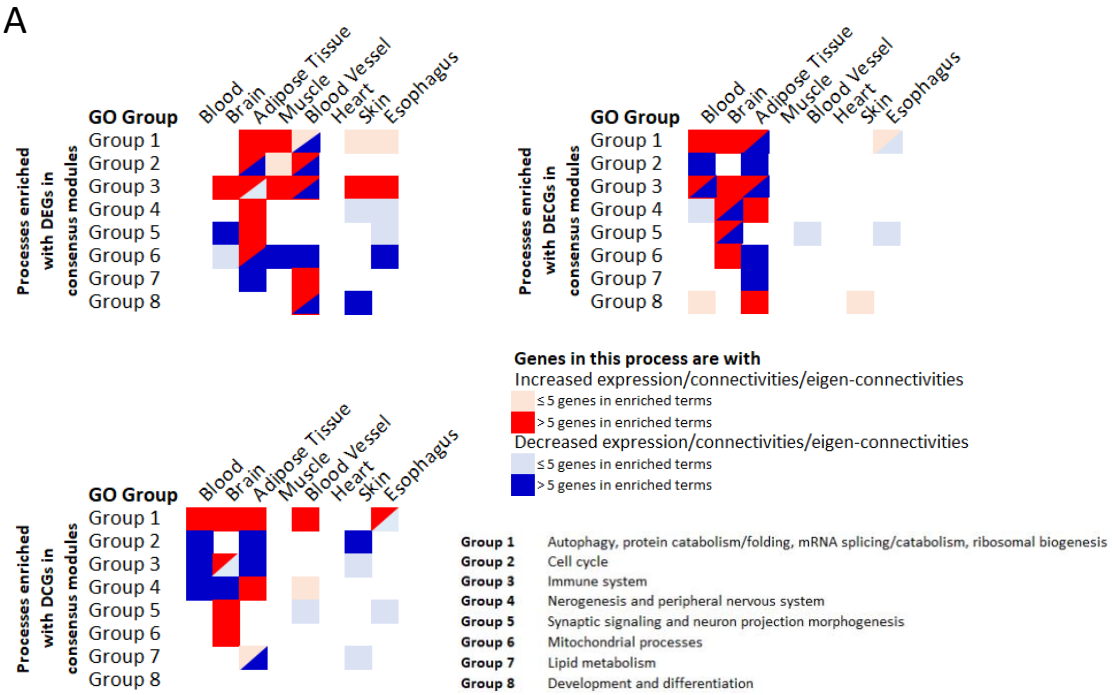
The overlaps between the genes found in the three analyses shows that only a small proportion of genes were identified simultaneously in the centrality analyses and in the DEGs analysis (Figure 2B). The tissues with most overlapping genes detected in the three metrics were adipose tissue and blood vessel. These tissues had a considerable number of DEGs, totaling roughly a third of all analyzed genes. However, even in these tissues, the fact that the overlap between DEGs and DCGs/DECGs was less than half of the total genes detected supports the conclusion that absolute differential expression and differential connection analyses are measuring distinct biological phenomena. On the other hand, DCGs and DECGs seem to be measuring more similar changes.

Consensus network and GO enrichment analysis of age-related sub-modules: Since we found that the connectivity analyses identified different novel sets of altered genes with aging, we set out to ask whether these represented biologically relevant processes looking for enriched functional terms. To get a better understanding of what types of biological processes are being modulated with age we clustered the genes into modules using WGCNA. We built an unsigned consensus network for the nine datasets in each tissue, using the same soft-threshold powers used before. This allowed the identification of conserved gene co-expression modules that could be found in all nine datasets and that, therefore, were stable across the adult lifetime. These modules represent genes that have high co-expression and high structural overlap, and usually represent meaningful biological modules that yield meaningful enriched terms when analyzed. After obtaining these modules, we overlapped them with the genes found in our previous analyses. This led to the definition of sub-modules, where each sub-module is comprised of the DEGs, DCGs, or DECGs that are contained in that module. By dividing the DEGs/DCGs/DECGs using the modular structure provided by WCGNA, we can group them into sets that are likely to be related to the same biological processes and, therefore, identified based on a functional enrichment analysis. This approach provides more detailed information about the enriched processes than the more commonly carried out method of enriching all altered genes in a single batch.

We enriched the DEGs/DCGs/DECGs containing sub-modules using topGO[28] in R. The enriched terms were grouped into sets that refer to processes that were recurrently found to be enriched. Some processes were frequently found to be enriched in the same modules, such as autophagy, RNA splicing, and ribosome biogenesis, suggesting that they work in a coordinated fashion. Thus, they were included in the same group. An extensive list of enrichment results per module can be found in Supplemental Data 2, 3, and 4, along with the full classification spreadsheet. All the classifications were done using terms that had at least two hits and were significant at an FDR threshold of 0.05, with either the classic or the weight01 algorithms in topGO. Most sub-modules were composed, entirely or almost entirely, of positive or negative DEGs/DCGs/DECGs, while only a few were more divided. To do the enrichment, we further divided the sub-modules into sets with genes with positive or negative slopes in each metric, enriching each set separately. Figure 3A summarizes the classification results for the sub-modules in each tissue,

for each of the three metrics. Groups presented here are only those found in at least two tissues. Additional tissue-specific enriched processes can be seen in the Supplemental Data 2, 3 and 4.

A



B

Heart - Group 1

Ensembl Gene ID	Symbol	Type	Description	Process	Slope
DEGs					
No Group 1 DEGs					
DCGs					
ENSG00000095917	TPSD1	protein_coding	tryptase delta 1	proteolysis	Positive
ENSG00000112183	RBM24	protein_coding	RNA binding motif protein 24	RNA splicing	Positive
ENSG00000203867	RBM20	protein_coding	RNA binding motif protein 20	RNA splicing	Positive
ENSG00000104412	EMC2	protein_coding	ER membrane protein complex subunit 2	protein insertion into ER membrane	Positive
DECGs					
ENSG00000104412	EMC2	protein_coding	ER membrane protein complex subunit 2	protein insertion into ER membrane	Positive

Figure 3 - GO enrichment of sub-modules derived from consensus modules analysis. A) Classification of the enrichment terms of sub-modules into processes across all tissues. Sub-modules are defined as the subset of genes in a particular module that is altered in one analyzed metric. DEGs/DCGs/DECGs with positive and negative slopes were enriched separately. B) List of Group 1-related genes in the three metrics in the heart.

Interestingly, some GO groups displayed relative consistency on whether their component genes have increased or decreased connectivities. Notably, Group 1 was recurrent and consistently positive across tissues and metrics. Group 3, on the other hand, although also very recurrent, displayed inconsistent behavior. While Group 3 DEGs are primarily positive, DECGs and mainly DCGs display more negative slopes. This analysis revealed no significant terms at FDR < 0.05 for heart. However, using less stringent FDR cutoffs, Group 1-related terms were detected, as shown in Figure 3B. Perhaps because heart had very few genes changing in all metrics, it was not possible to generate a list extensive enough to do a proper enrichment analysis. However, the fact that the most enriched terms, even if at FDR higher than 0.05, are related to Group 1, the most recurring group in all tissues, demonstrate that the processes in this group seem to be

consistently altered with age in all the analyzed tissues. It is worth noting that heart had 15 DCGs with positive slopes, four of which could be assigned to Group 1.

Comparisons of altered genes across tissues: While Figure 3 shows that some processes are being similarly changed across tissues, the question remains whether the genes involved in such changes are the same. To try to understand that we compared the DEGs, DCGs, and DECGs found in each tissue. Figure 4A displays the overlap between the five tissues with the most altered genes in each metric, and an overlap when we consider all three metrics combined (unique genes altered in at least one metric). Only five tissues are shown because visualizing more than five sets with a Venn diagram is difficult. A complete comparison of all eight tissues was carried out with the UpSetR R package[29] and all the results can be found in Supplemental Figures 1, 2, and 3. Figure 4A shows a minimal overlap between the five chosen tissues, despite the conserved processes observed in Figure 3. DEGs displayed more overlaps, with a higher proportion of DEGs present in at least two tissues (Figure 5A). DCGs and DECGs have a higher proportion of genes altered in a single tissue (Figures 4A and 5A).

Most of the genes that are shared in any of the tissues are only shared between two tissues, and only DEGs showed more than 10% of genes overlapping in at least three tissues (Figure 5A). Since Group 1 seemed to be recurrent across most tissues and metrics, we further analyzed modules in this category. We cross-tabulated the modules to try to understand whether their genes were similar. We first cross-tabulated all the modules and then sub-set all the modules classified as Group 1 in all the tissues. Supplemental Data 5 shows the result of this cross-tabulation, with corresponding overlap p-values for each module pair. The overlaps correspond to the whole modules and not only to the DEGs/DCGs/DECGs containing sub-modules. Many module pairs present a high overlap with very significant p-values. Out of the sub-set Group 1 modules, we identified a cluster of module pairs that seems to have high overlap with each other across most tissues: module 11 in the brain, module 21 in adipose tissue, module 6 in muscle, module 15 in the blood vessel, module 4 in heart, module 15 in skin and module 8 in the esophagus (Figure 4B). In blood, the overlapping module was the vast module 1, which makes its use in the comparison difficult. However, we have included it in a few comparisons. We then overlapped the genes in these selected modules using two approaches. In the first, we overlapped the genes in the whole module. In the second, we overlapped only the DEGs/DCGs/DECGs in the sub-modules. Here, again, we used, at most, five tissues for each metric to compare in the Venn diagram, or fewer when less than five tissues had a reasonable number of genes altered. The left panel in Figure 4C shows that when overlapping the whole modules, there is much overlap between these modules. Sixty-three DEGs are shared between the selected modules in five tissues, 66 DCGs are shared between four tissues, and 66 DECGs between three tissues, each representing roughly half of the modules, except when the larger brain module is used. Other overlaps are also shared between triplets. This shows that despite the modules being found independently in each tissue, they captured clusters of genes that behave in a coordinated fashion in most tissues. However, when overlapping DEGs/DCGs/DECGs containing sub-modules, there is a remarkable lack of overlap, with the vast majority of DEGs/DCGs/DECGs being either unique to the tissue or shared with only one other tissue.

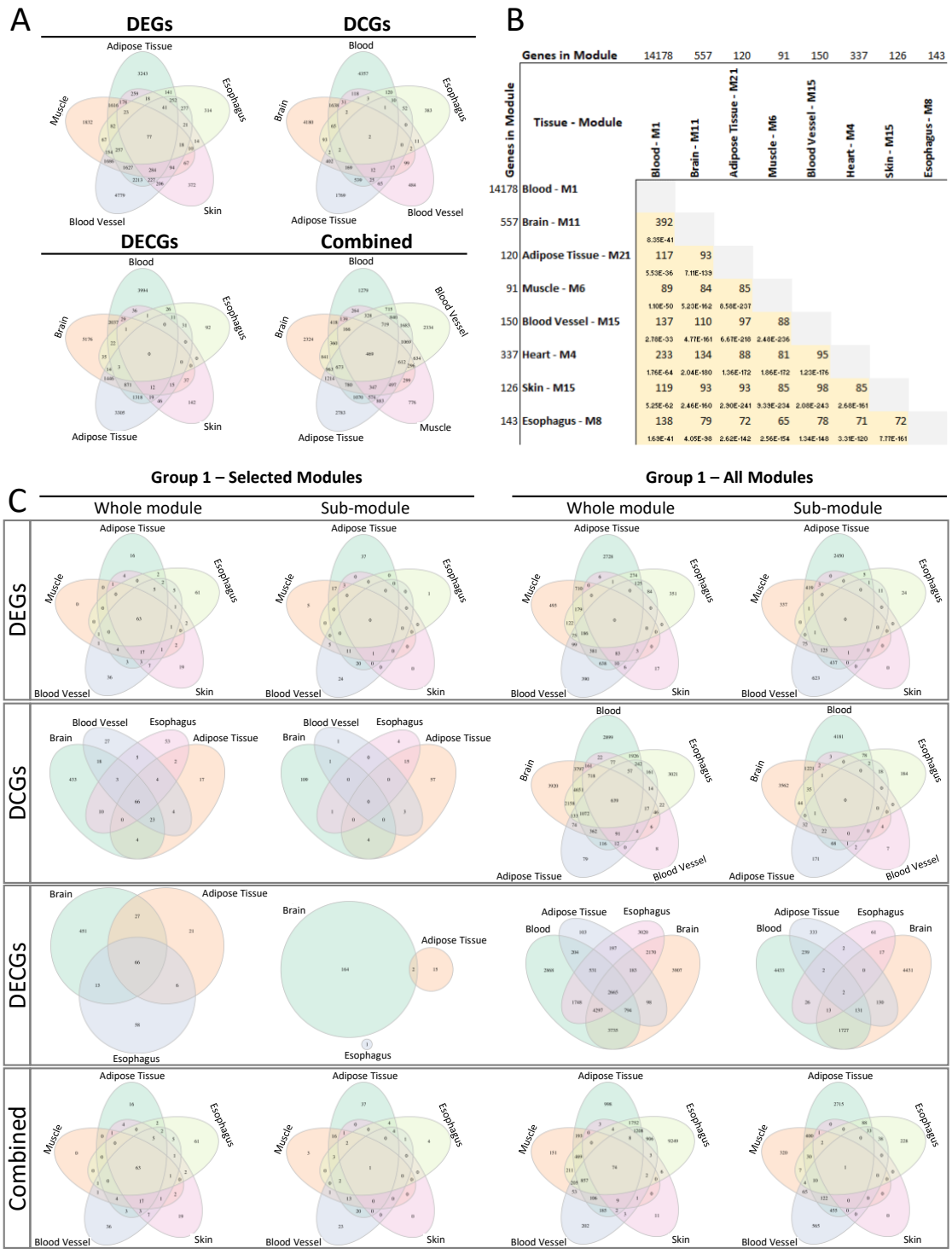


Figure 4 – Evaluation of the conservation of altered genes across tissues. A) All the DEGs, DCGs and DECGs in the five tissues with the most altered genes in each metric were overlapped to assess their conservation across tissues. B) Cross-tabulation of Group 1 modules with high overlap with each other; the full cross-tabulation table is presented at Supplemental Data 5. The number in each table entry represents the total number of genes overlapped in each module pair; the small number represents the overlap p-value. Table entries with a yellow shade both here and in Supplemental Data 5 have an overlap p-value < 0.05. Since the table here was subsetted for modules with high overlap, they are all shaded yellow. C) Venn diagrams of Group 1 modules in different tissues. The first two columns on the left, under "Group 1 – Selected Modules", show the results considering only the selected Group 1 modules with high overlap shown in Figure 4B. Columns on the right, under "Group 1 – All Modules", show results considering all modules that

were classified as Group 1. For each of these two sets, we overlapped both the full set of genes inside these modules (including non-DEGs/DCGs/DECGs), or only the set of genes in the sub-modules (only DEGs/DCGs/DECGs inside the modules). A comparison with combined unique DEGs/DCGs/DECGs was also provided in the last row. The number of tissues overlapped in each scenario was at least three, and at most 5, depending on the tissue having a large enough number of genes in that scenario for comparison.

We then made the same comparison using all the modules and sub-modules classified under Group 1. The right panel in Figure 4C shows the results obtained, in the same fashion as the comparison with only the selected modules. It is worth noting that despite showing no overlap between the five selected tissues, the result for DEGs has an overlap of 186 DEGs when skin is not considered. Skin only produced Group 1 modules concerning DEGs, probably decreasing the likelihood of an overlap when compared to the other tissues, which displayed much more robust results. The combined metrics results are also displayed in Figure 4C, demonstrating that even when considering genes altered in any of the three analyses, their overlap is minimal. These results show that despite the same modules/processes being affected, as shown in Figure 3, different targets inside these modules/processes are changing with age in each tissue.

To determine to which process these overlapped genes belong, we enriched all the gene sets present in more than one tissue, as shown in Figure 5A (Supplemental Data 6). The classification of the enrichment results is presented in Figure 5B. The results show that overlapped genes in the three metrics tended to fall into Group 1, despite these results being obtained only in the sets of genes overlapped in 2 or 3 tissues. Group 3 was also observed in the DEGs overlapping five tissues.

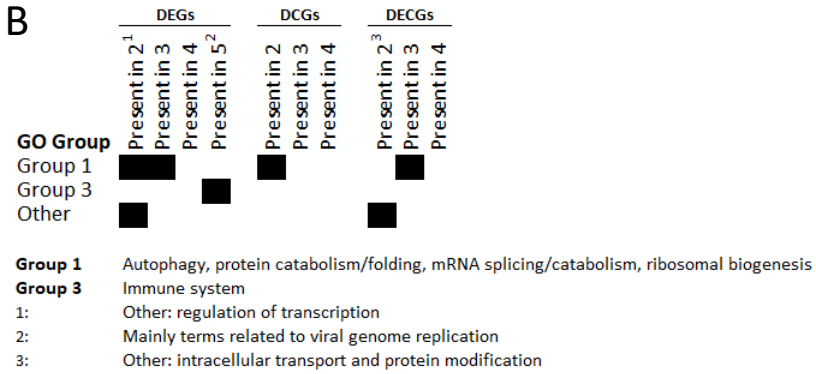
A

DEGs		
Present In	**N° of Genes**	**Proportion of Total**
1 tissue	10653	51.44%
2 tissues	6609	31.92%
3 tissues	2720	13.14%
4 tissues	633	3.06%
5 tissues	86	0.42%
6 tissues	5	0.02%
7 tissues	2	0.01%
8 tissues	0	0.00%
Total	**20708**	**100%**

DCGs		
Present In	**N° of Genes**	**Proportion of Total**
1 tissue	11213	75.84%
2 tissues	3202	21.66%
3 tissues	351	2.37%
4 tissues	18	0.12%
5 tissues	2	0.01%
6 tissues	0	0.00%
7 tissues	0	0.00%
8 tissues	0	0.00%
Total	**14786**	**100%**

DECGs		
Present In	**N° of Genes**	**Proportion of Total**
1 tissue	12712	67.85%
2 tissues	5018	26.78%
3 tissues	990	5.28%
4 tissues	16	0.09%
5 tissues	0	0.00%
6 tissues	0	0.00%
7 tissues	0	0.00%
8 tissues	0	0.00%
Total	**18736**	**100%**

B



Group 1 Autophagy, protein catabolism/folding, mRNA splicing/catabolism, ribosomal biogenesis
Group 3 Immune system
1: Other: regulation of transcription
2: Mainly terms related to viral genome replication
3: Other: intracellular transport and protein modification

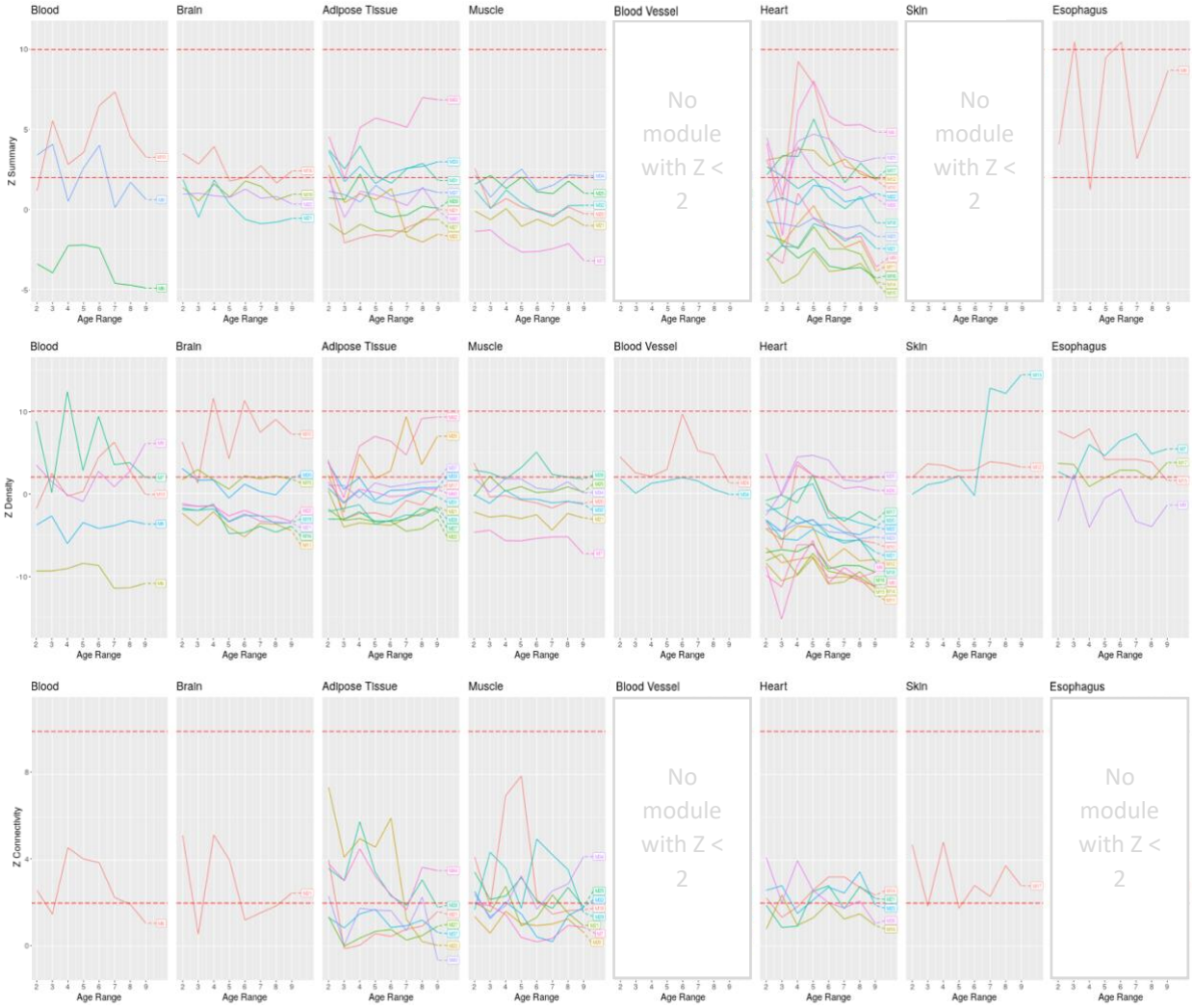
Figure 5 – Intersect count and enrichment. A) Gene count of DEGs, DCGs, and DECGs detected in sets of 1 up to 8 tissues. The tables display the number of genes that were altered in a specific number of tissues and the proportion that these altered genes represent concerning the total amount of altered genes from the combined set of the three metrics. B) Classification of GO enrichment results of the intersection sets.

Module preservation analysis: The age-associated gene modules identified through the consensus network analysis described here identified are the most conservative ones, since each module must be present in all datasets to be detected. However, even if we can define more general clusters of genes that are conserved throughout the datasets, this does not mean that the network structure is conserved with age.

To ask whether network modules are conserved with age, we used the module preservation statistics defined by Langfelder et al. (2011)[30]. In this analysis, several network preservation statistics are calculated for each module to compare a given network to a reference network. The reference networks used in this analysis were the networks of the youngest age ranges in each dataset. Therefore, we identified modules using only the youngest network in each tissue. Then, the gene-module labels/colors defined by this first network were superimposed on the networks of the subsequent age ranges. A series of network statistics were then calculated to try and provide evidence of whether a specific module found in the youngest network can be said to be preserved in the older networks. For that, two sets of preservation measurements are calculated: density-related calculations, which ask whether the group of genes defined by the module present a denser connection with each other than with the background; and connectivity related calculations, which ask whether the hub genes' statuses are preserved in the reference and test networks. The background to which these calculations are tested against is defined by a random sample of genes, and the statistics is measured by a permutation Z score. These statistics are described in more detail in the methods section.

These calculations were done for all modules in all tissues. Since there are nine age ranges, and the first one is used as the reference network, each module in each tissue is tested eight times. Figure 6A presents the network preservation statistics of all eight comparisons for each module in each tissue. The figure shows only modules for which the module preservation statistics failed to show evidence of preservation (Z score < 2) in at least one age range. Z scores for all modules can be found in Supplemental Data 7. Density-related measurements are summarized by the Z density score, whereas the Z connectivity score summarizes connectivity-related measurements. Z summary aggregates both Z density and Z connectivity. Several modules showed evidence of non-preservation in at least one age range, while others showed weak evidence of non-preservation in one or two age ranges, just to recover evidence of preservation in other age ranges. For instance, brain's module 10 Z density falls slightly below 2 in just one age range, so that overall, this module may be considered preserved. However, since there is no meaningful way of establishing a criterium of how many times and at which age ranges a module's Z score can fall below 2, we considered for further analysis any modules found to be non-preserved in least one point.

A



B

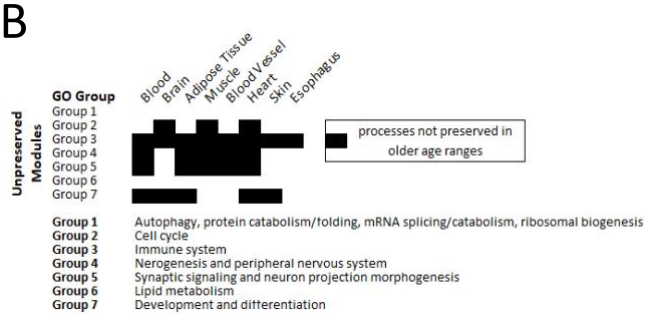


Figure 6 - Module preservation analysis using the network from the youngest age range as reference. A) Z scores of unpreserved modules with age range. Only modules with Z score (Z summary, Z density, or Z connectivity) lower than two in at least one age range are included. B) Summary of enriched GO terms for unpreserved modules in each tissue. Enrichment was carried out for all modules with a Z score lower than 2 in any of the three scores (Z summary, Z density, or Z connectivity). Only GO terms with FDR < 0.05 were considered for the classification.

However, some modules showed a more consistent pattern of losing preservation over time, like adipose tissue's module 22 Z connectivity, falls below 2 at age range 7 and remains non-

preserved. Since some preservation statistics can be influenced by module size, an alternative median rank statistic was calculated (Supplemental Figure 4), which typically shows no dependence on module size. Worst-ranked modules, according to these statistics, tend to recapitulate the worst preserved modules in Figure 6A. Interestingly, heart showed a high number of modules that seemed unpreserved, primarily by density criteria at very young age ranges. In the differential connectivity and expression analyses shown in Figure 2, heart was the only tissue with a low number of genes altered by using the three metrics. All the other tissues show at least a thousand altered genes in at least one of the measurements.

Next, we carried out a GO enrichment analysis of the unpreserved modules to better understand what types of processes were being affected. Figure 6B summarizes biological processes found in this manner. One interesting observation is that the same types of processes that were used to categorize the enrichment in the consensus network analysis could also be used to categorize the results of this analysis.

To better characterize the relationship between the differential expression and connectivity results and the module preservation analysis, we overlapped the DEGs/DCGs/DECGs in each tissue, clustered according to positive or negative slopes, with the modules found in the reference networks in those tissues. Figure 7A shows the modules in the reference (youngest) network in each tissue ordered by the percentage of genes contained in each that are positive (top panel) or negative (bottom panel) DEGs/DCGs/DECGs. Bars highlighted by the bordering rectangle indicate the unpreserved modules found in Figure 6A. As noted in the top panel, modules with a high proportion of positive DCGs (at the top of the ordered chart) were not found to be unpreserved in any age range. Since positive DCGs are having the ranked sum of their edges increased, this is probably reflected in denser connections with time inside their modules, or at least preserved. Furthermore, since the most prevalent types of processes that seem to have a consistently increased connectivity are the ones in Group 1, it is possible that the maintenance, and maybe even strengthening, of the connections between the genes in these processes is essential during aging.

Interestingly, unpreserved modules seem to be more evenly distributed among the modules with different proportions of negative DCG content, with some of the top-ranked modules by negative DCG content showing evidence of non-preservation (Fig. 7A, bottom panel). However, none of the negative DCG proportions were significantly high, suggesting that the decrease in the ranked sum of edge weights in negative DCGs may facilitate the breakdown of modules. Concerning DEGs and DECGs, these relationships also seem to exist but are less apparent. While for DCGs no unpreserved module had more than 25% positive DCG content, for DEGs and DECGs two unpreserved modules had more than that (adipose tissue module 42 and muscle module 25). The two top modules in terms of positive DEG and DECG content that were unpreserved in adipose tissue (modules 42 and 20) had very little DCG content, likewise for muscle module 25, which has no DCGs. Therefore, the DCG content may be more relevant to understanding the biological phenomenon. Likewise, with DCGs, the negative content of DEGs and DECGs was never high, and unpreserved modules seem to be more evenly distributed (Fig. 7A, bottom panel).

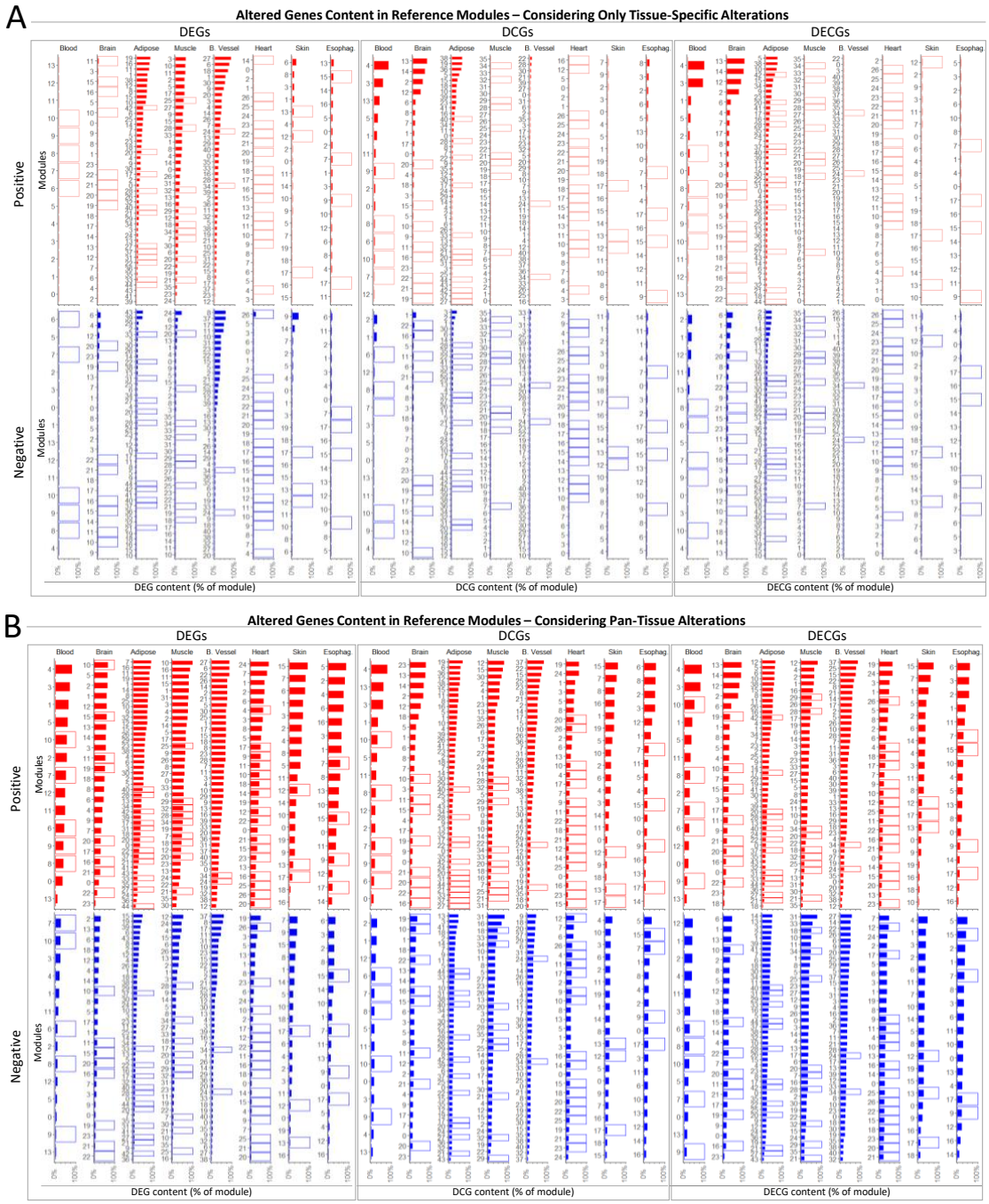


Figure 7 – Overlap of positive and negative DEGs/DCGs/DECGs with the modules of the reference network. A) Proportion of tissue specific DEGs/DCGs/DECGs with positive (top panel) or negative slopes (bottom panel) in each module of the reference age range. B) Proportion of DEGs/DCGs/DECGs with positive (top panel) or negative slopes (bottom panel) in each module of the reference age range, combining all DEGs/DCGs/DECGs from all tissues, and not only those altered in the tissue in question. The highlighted bars represent modules that were shown not to be preserved according to Figure 6A. In both panels, the highlighted bars represent modules that were found not to be preserved, as shown in Figure 6A.

To test the hypothesis that modules that are typically affected in aging by increasing the centrality of their component genes are more likely to be preserved, independently of the actual increase in centralities, as they would represent more biologically cohesive modules, we used the same strategy of overlapping the DEGs/DCGs/DECGs with the reference network's modules, but instead of overlapping only DEGs/DCGs/DECGs that changed in that tissue, we combined the results of all tissues. We overlapped the modules within a set with all the positive DEGs/DCGs/DECGs in all tissues and another set with the negative ones, to see whether modules that contained genes that typically would be differentially expressed or connected in each environment (tissue) with aging are related to the preservation/non-preservation status, irrespective of the fact that they were differentially expressed or connected in that tissue. Figure 7B shows the results of this analysis. Like what we found in the tissue-specific analysis (Figure 7A) the top-ranked modules by positive DCG content in all tissues show no indication of non-preservation. In contrast, non-preservation was found among the top-ranked modules by positive DEG and DECG content. However, all the top-ranked modules by positive DEG/DECG content that were non-preserved were bottom-ranked in the positive DCG content, supporting the idea that positive DCG content is negatively associated to whether a module will not be preserved with aging. The fact that this relationship held even when using DCGs that were not tissue-specific suggests that there may be some feature of the genes in these modules that preserve their relationship during aging, irrespective of the actual positive centrality change observed in any one tissue.

Discussion

With few notable exceptions, aging is a universal phenomenon. But despite significant advances in finding molecular pathways that may control the aging process, it is still unclear whether these pathways are conserved among different species, or even among different tissues in the same individual. In that regard, looking for conserved patterns in biological endpoints has been extensively used to try to find such conserved pathways. Using gene expression data as biological endpoint, Zhan et al. (2006)[11] found 250 DEGs with age, enriched in symporter activity, sialyltransferase activity, and chloride ion transport related terms in human muscle. However, when comparing their results with brain and kidney DEGs from other studies[31; 32], they observed minimal overlap. Similar findings were reported by Glass et al. (2013)[12], analyzing skin, adipose tissue and lymphoblastoid cell lines; and by Voutetakis et al. (2015)[13], analyzing skeletal muscle, and nervous tissue. These studies revealed tissue specific DEGs with age, enriched in specific biological pathways, but with little overlap between each other. Moreover, Yang et al. (2015)[14] analyzed multiple tissues from version 3 of the GTEx dataset and also found little similarity, with the most common terms annotated for downregulated genes being mitochondria-related, in five out of seven tissues (adipose tissue, artery, heart, lung, and blood). More recently, Palmer et al. (2021)[33] performed a meta-analysis of 127 microarray and RNA-Seq datasets from humans, mice, and rats, which included brain, heart, and muscle tissue and also found limited overlap between genes altered in different tissues. They did, however, observe a few similarities in enriched GO terms, such as decreased expression of mitochondrial genes and increased immune response and inflammation-related genes. Other studies employed WGCNA to explore age-related changes. Hu et al. (2018)[17] identified modules significantly associated with age in the human prefrontal cortex, finding processes related to transport, cell signaling, mitochondrial organization, and nitrogen compound metabolism. They did not, however, carry out a comparison with other tissues. Ferreira et al. (2021)[20] applied WGCNA integrated with non-linear differential expression analysis in mouse tissues, and identified modules associated with immune system (in brain and liver), mitochondrial processes (in heart), and

development/differentiation (in muscle). Nonetheless, immune system was the only process common to at least two tissues.

The lack of common pathways differentially expressed with age in most tissues could be interpreted as if, mechanistically, the aging process is different in each tissue of an organism. However, we hypothesized that functional changes, such as those seen in aging, may not be due solely to differences in absolute gene expression levels, but also to differences in how they co-express with others in a systems level. Thus, we have combined analysis of gene expression changes to analysis of changes in co-expression by using network topological metrics (centrality, eigenvector centrality, as well as several module preservation statistics). This approach yielded very different results from a traditional differential expression analysis, revealing novel patterns of transcriptomic changes that cannot be seen by looking only at absolute gene expression values, as some tissues with very few DEGs displayed many DCGs and DECGs. In fact, apart from adipose tissue, all other tissues analyzed here displayed an inverse relationship between differential expression and connectivity. The gene sets that resulted from the two connectivity analyses (DCGs and DECGs) had a good overlap with each other, although still with substantial differences. However, both measurements had poor overlap with the DEGs found for the same tissues.

In addition, while most studies usually enriched the full set of DEGs at once, we opted to cluster the genes into modules to facilitate identifying which processes were changed, since more general, and hence common, terms may obscure specific pathways when all genes are enriched at once. Therefore, by clustering genes according to their co-expression and topological overlap, we create sets of functionally related genes. Then, by overlapping these modules with the sets of DEGs, DCGs, and DECGs, we can derive sub-modules that, when enriched, provide a more accurate reading of which processes are changing with aging. Some other articles have worked with modules in aging [17; 20] but using different approaches. For instance, by using the overlap of three sets (module genes, genes derived from non-linear differential expression analysis, and hub genes), Ferreira et al. (2021) ended up with sets composed of a very small number of genes, possibly limiting the power of enrichment analysis. Our approach resulted in several enriched terms present across tissues. Moreover, by clustering genes, we can see which genes have correlated changes and which terms seem to be correlated. For example, the same modules showed up in terms related to autophagy, protein catabolism and folding, mRNA splicing and catabolism, and ribosomal biogenesis, indicating that these processes are altered with age and that they themselves have a relationship with each other. Hence, we classified the modules with enriched terms into groups, from which eight were chosen for further analysis as they appeared recurrently. Other non-recurrent enriched terms can also be found in the full enrichment results in the supplemental material.

Our results show that Group 1 terms are enriched in almost every tissue according to at least one metric, with most genes being upregulated with age. Interestingly, even when Group 1 modules were not revealed using one metric, they were in another. For example, in muscle DCGs and DECGs were not identified, and no muscle Group 1 module appeared. On the other hand, muscle DEGs were identified, as were Group 1 modules for this metric. Similarly, the small number of blood DEGs did not yield any Group 1 module. In contrast, Group 1 modules with several genes were found with DCGs and DECGs. When all metrics are considered, Group 1 enrichment can be found in all tissues except the heart. However, heart was the only tissue with very few altered genes in all three metrics. Nonetheless, when these heart genes are manually inspected, several Group 1 genes are found among those upregulated with age (Figure 2B).

Group 3 genes, related to the immune system, are also recurrently found although, in this case, results were not consistent across different metrics. For DEGs, Group 3 genes were consistently found to be upregulated with age. On the other hand, for DCGs, Group 3 showed signs of decreased centrality in several tissues but increased in the brain. Results with DECGs are even more mixed. While changes in expression are more straightforward to interpret, changes in centrality are less intuitive. A decrease in centrality in this situation means a decrease in the ranked sum of the correlations of the gene with every other gene. The gene is, therefore, less co-expressed overall when other genes are expressed. Due to the ranking of the centralities during the analysis, we cannot establish whether the centrality decreased or remained the same while its rank decreased due to others' centralities increasing. If the first alternative is the case, this decrease in the centrality of immune system-related genes could be interpreted as that even though the overall expression of immune system genes is increased with age, the process as a whole is dysregulated, with the genes not co-expressing as strongly with the same partners as the tissue ages.

Group 6 genes, related to mitochondrial processes, were also recurrent. DEGs were mostly downregulated, in line with previous studies. Few tissues, however, had Group 6 modules identified within DCGs and DECGs, with the brain being the only one that displays Group 6 modules in all metrics. Overall, our results clearly show that several biological processes display transcriptomic alterations with age across tissues in a conserved fashion, but that these changes seem to manifest in different ways in different tissues, such that simple differential expression analysis cannot detect them in some tissues. The analysis of co-expression variations, in the form of differential connectivity analysis, however, is capable of revealing alterations when only differential expression is not.

While the lack of overlap of individual genes altered with age in different tissues had been found before, our results show that this lack of overlap is, in fact, remarkable. Figure 4A shows that most genes altered with age are present in only one tissue. A considerable proportion of the remaining is present in only two tissues (Figure 5A). Interestingly, the overlaps are more significant when the three metrics are combined. The analysis of Group 1 modules (Figures 4B and C) clearly demonstrates the lack of overlap. Even when modules with high overlap are compared, the genes being altered within these modules are not the same. When overlapping sets of genes are enriched (Figure 5B), they reveal mostly Group 1-related terms, indicating that some Group 1-related genes are changing in different tissues but that most Group 1-related genes are changing in a tissue-specific manner. This observation illustrates the difficulties in finding similarities between tissues when looking at individual genes. Nonetheless, even though the set of same genes changing in several tissues is small, changes in the same processes can be observed, which could be interpreted in light of the fact that different tissues have specific needs and rely on the same pathways differently. Therefore, the same alteration/damage in different tissues may manifest in slightly different ways.

The module preservation analysis used here also helped shed light on the changes in network structure that happen with age. First, when non-preserved modules are enriched, we find the same groups as revealed in the other analyses. But interestingly, there is a notable lack of Group 1 and Group 6 modules. Given that these two groups were the most prevalent ones in the other analyses, as well as the most consistently upregulated across tissues, it is curious that they were precisely the ones not found in the module preservation analysis in any tissue. This indicates that the relationship between the genes related to these processes remains stable, or may even be strengthened, with aging. Group 3, on the other hand, was altered in all tissues, indicating that immune system-related modules seem to have a more flexible clustering pattern with age, a possible sign of degradation in the coordination of their component genes, or an adaption to different (patho)physiological needs as aging progresses. Even in heart, for which the results of

the GO enrichment of sub-modules derived from consensus modules (Fig 3A) does reveal changes in any group, there was evidence for non-preservation in this analysis in similar groups. Therefore, we believe these two network analyses complement each other and suggest that relationships between some biological processes may become strengthened, or at least better preserved, with age, while others have their network structure degraded.

By overlapping the reference network in the module preservation analysis with the DEGs, DCGs, and DECGs, we could explore whether DCG content was related to the preservation of modules. Overall, the results indicate that modules with genes that are typically positive DCGs tend to be preserved, even if those genes are not actually being differentially connected in that specific tissue.

Of all the groups found in our analyses, Group 1 is noteworthy. While Group 3 (immune system) and Group 6 (mitochondrial processes) reflect processes that have already been mechanistically implicated in aging research[34; 35], Group 1 i) seems to have a more heterogeneous set of enriched terms, and ii) these enriched terms can be very easily related to the life span extension promoted by mTORC1 inhibition. Inhibition of mTORC1, through genetic or pharmacological means, is one of the most studied methods of life span extension and works in several organisms, including mice[36-51]. Therefore, higher expression and relative centrality of autophagy and protein catabolism, as indicated in the Group 1 modules, demonstrate that these mTORC1-related processes are more prominent during aging. Also, in Group 1, we find mRNA splicing and protein folding genes. Loss of proteostasis is a hallmark of aging[2], and aberrant splicing is a cause of malformed proteins, which demand more of the protein folding mechanisms in order to try to correct these malformations. Aging seems to be accompanied by a high incidence of aberrant splicing and intron retentions in *C. elegans*, fruit flies, mice, and humans[52-56]. Notably, Mariotti et al. (2022)[56] showed an increase in these events in the GTEx dataset for several tissues. Interestingly, in their analysis brain presented the least significant association between intron retention and age. Curiously, our analysis revealed no Group 1-related DEGs in brain, but abundant Group 1 DCGs and DECGs, which might suggest a different relationship between splicing and aging than in other tissues. Blood also displayed Group 1 DCGs and DECGs, and no DEGs, but blood had a minimal amount of DEGs (19), contrary to the brain that had hundreds of DEGs.

Ribosomal biogenesis has also been associated with aging[57]. Curiously, previous studies have pointed to the downregulation of ribosomal biogenesis with age[57], contrary to the increase we observe here. Nonetheless, the presence of these processes in the same modules suggests that they work in a coordinated fashion. Higher rates of splicing abnormalities may lead to more malformed proteins, which will demand more of the protein folding machinery. Such aberrant mRNAs may create difficulties for the ribosomal machinery since they may take longer to translate completely, given the difficulties in folding. This increased occupancy of ribosomes would end up requiring the biogenesis of more ribosomes, to decrease the amount of mRNA in line to be translated. To counter this increase in aberrant mRNAs, proteins, and eventually organelles negatively affected by malfunctioning proteins, catabolic processes such as RNA and protein catabolism, and autophagy may be more demanded. Since these clearance mechanisms seem to be upregulated with age, mTOR inhibition may be exerting lifespan extension by boosting a naturally occurring adaptation mechanism. Perhaps the fact that these processes, as a group, are more demanded with age can explain why they were not found to be not preserved in the module preservation analysis, as proper coordination of these processes is more required with age.

The fact that such alteration is not seen in blood and brain with DEGs, but is with DCGs and DECGs, suggests that such relationships may be more intricate than simple absolute levels of expression. First, differential expression analysis is usually carried out in normalized samples (in our case, TPM). Therefore, the values represent a relative change in gene expression and not actual absolute values. The expression of genes in the Group 1 processes may be increasing in all tissues, even if their relative proportion remains the same. Another possibility is that expression levels do not change, but differences in network topology measures reflect differences in signaling or in the proportion of different transcripts for the same gene. Post-translational modifications could make the same proteins induce different processes, leading to different gene co-expression outcomes. Likewise, if a different isoform of a given protein becomes proportionally more expressed, this could change the co-expression relationships between that gene and others.

It is certain that further work is required to dissect the molecular mechanisms that are reflected in changes in connectivity and module preservation. However, our work clearly shows that relying only on simple differential expression analysis can conceal changes in pathways that may be more relevant for understanding the aging process, or any other biological process that is mechanistically associated with changes in gene expression. Importantly, it should be emphasized that gene expression data can be analyzed using different metrics to measure co-expression patterns, allowing the identification of different kinds of relationships related to the same biological processes. Taken together, the combined analyses we carried out here using transcriptomic data show that, even though there are some tissue-specific alterations, some processes seem to be changing across all tissues, even if with different strengths and manifested with different dynamics, supporting the idea that the aging phenotypes are associated with alterations in a few, conserved, molecular pathways.

Methods

Dataset: Gene expression data were obtained from the Genotype-Tissue Expression (GTEx) project. GTEx (v8 release) has postmortem RNA-seq data from 948 donors, aged 20 to 70 years old, collected from 54 tissues, totaling 17,382 samples. To increase the power of the analyses and allow the division of the dataset into age ranges, with an adequate number of samples in each age range, we selected only tissues with at least 800 samples, such that were analyzed data from blood, brain, adipose tissue, muscle, blood vessel, heart, skin, and esophagus. Tissues were analyzed at the level defined by the SMTS sample attribute in the GTEx annotation files. Therefore, some tissues analyzed here are comprised of aggregated sub-tissues.

Detection of differentially expressed genes with linear mixed model: Differentially expressed genes were detected for each tissue separately using the following linear mixed model:

$$Y_{ijk} = \beta_{0j} + \beta_{1j}Age_i + \beta_{2j}Sex_i + \beta_{3j}Race_i + \beta_{4j}Ischemia_i + \beta_{5j}Batch_i + \beta_{6j}Death_i + \beta_{7j}Individual_k + \epsilon_{ijk}$$

where, Y_{ijk} is the expression level (log2 normalized TPM + 1) of gene *j* in sample *i* of individual *k*, Age_i is the age of sample *i*, Sex_i is the sex of sample i, $Race_i$ is the race of sample *i*, $Ischemia_i$ is the time of ischemia of sample *i*, $Batch_i$ is the experimental batch in which sample *i*'s gene expression was measured, $Death_i$ is the type of death, in the Hardy scale, of the donor from which sample *i* was collected, and ε_{ijk} is the error term (assumed to be normally and independently distributed with var(ε) = σ^2I). Individuals are considered a block random effect,

considering that some tissues have more than one sample for the same individual. Ischemic time was divided into 300-minute intervals to standardize the procedure with the removal of confounding factors, done in a later step. Random effects ($Individual_k$) are assumed to be normally and independently distributed with var(*Individual*) = $\sigma_{ind}^2 I$.

Genes with no or small variation in expression were filtered out of the analysis. Selected genes had expression higher than 0.1 TPM in at least 20% of the samples. To be considered differentially expressed with age, genes needed to satisfy an FDR<0.05, considering BH multiple test corrections. The model was fit using the nlme package in R[22].

Removal of confounding factors: Expression was corrected by confounding factors using ComBat, from the sva R package[24]. The correction was made for sex, race, experimental batch, ischemic time, and death type. Given that ComBat does not accept continuous variables, ischemic time was divided into 300-minute intervals. Since ComBat can only correct for one effect at a time, effects were removed iteratively, following the order presented previously in this paragraph.

Differential connectivity analysis: The dataset was divided into nine 6-year age ranges, except for the last one, with a 3-year range. This division ensures that the smallest sub-dataset has at least 20 samples, the minimum recommended for WGCNA. Connectivity (degree centrality, node strength) and eigenvector centrality measurements for each gene were then calculated in each age range using the *igraph* R package[26]. For comparison between networks, the connectivities inside each age range were ranked and a linear regression was applied to detect differences in ranked connectivities along the nine age ranges in the two metrics.

Co-expression network construction and consensus module identification with weighted gene co-expression network analysis: Gene expression data for each analyzed tissue and age range without the effects of known confounding factors were used as inputs for the WGCNA package in R[25]. WGCNA constructs weighted gene co-expression networks using a soft-threshold and fitting the data into a scale-free topology model, which is typical of biological networks. The chosen soft-thresholding powers for each tissue were: 6 (blood), 6 (brain), 9 (adipose tissue), 12 (muscle), 8 (blood vessel), 12 (heart), 6 (skin), and 6 (esophagus). We used the consensusBlockwiseModules function for module detection, with networkType set as "unsigned". All parameters were kept at their default values. Sub-modules were defined by overlapping the sets of genes of the three metrics (DEGs, DCGs, and DECGs) with the modules defined by consensus modules analysis.

Functional enrichment analysis: All GO enrichment analyses were carried out with the topGO R package[28], according to the documentation. The subsets of the sub-modules with positive and negative DEGs/DCGs/DECGs were enriched separately. Multiple test correction was performed using the BH method. Enrichment results were then classified into the appropriate groups. All the classifications were done using terms with FDR < 0.05 for the classic algorithm and p-value < 0.05 for the weight01 algorithm. Only terms with at least two significant genes were considered for classification.

Gene overlaps: Gene overlaps, and overlap significance statistics were calculated using the GeneOverlap R package[58].

Module preservation analysis: Module preservation analysis was carried out using the WGCNA R package[25]. The reference network was defined as the youngest network. The preservation of the modules found was evaluated through the network preservation statistics defined by

Langfelder et al. (2011)[30]. maxModuleSize was set to 1000 and nPermutations to 200. The minimum Z-score to consider a module preserved was set to 2, which is equivalent to a p-value of 0.05.

Acknowledgements: This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) grant # 2017/04372-0 to NCS-P. CMPFB was supported by Conselho Nacional de Pesquisa Desenvolvimento Científico e Tecnológico (CNPq) grant # 140207/2018-0.

CRediT authorship contribution statement: NCS-P, CMPFB and AF conceived and designed the project. CMPFB performed the analyses. NCS-P, CMPFB and AF wrote the paper.

References

1. Finch C. 1990. Longevity, senescence, and the genome. Chicago: University of Chicago Press.

2. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. 2013. The hallmarks of aging. Cell. 153(6):1194-1217.

3. Gems D, de Magalhães JP. 2021. The hoverfly and the wasp: A critique of the hallmarks of aging as a paradigm. Ageing Res Rev. 70:101407.

4. Bocklandt S, Lin W, Sehl ME, Sánchez FJ, Sinsheimer JS, Horvath S, Vilain E. 2011. Epigenetic predictor of age. PLoS One. 6(6):e14821.

5. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, Klotzle B, Bibikova M, Fan JB, Gao Y et al. 2013. Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol Cell. 49(2):359-367.

6. Horvath S. 2013. Dna methylation age of human tissues and cell types. Genome Biol. 14(10):R115.

7. Koch CM, Wagner W. 2011. Epigenetic-aging-signature to determine age in different tissues. Aging (Albany NY). 3(10):1018-1027.

8. Weidner CI, Lin Q, Koch CM, Eisele L, Beier F, Ziegler P, Bauerschlag DO, Jöckel KH, Erbel R, Mühleisen TW et al. 2014. Aging of blood can be tracked by dna methylation changes at just three cpg sites. Genome Biol. 15(2):R24.

9. Ake Lu VP, Fei Z, Raj K, Horvath S. 2021. Universal dna methylation age across mammalian tissues. Innov aging. © The Author(s) 2021. Published by Oxford University Press on behalf of The Gerontological Society of America. p. 410.

10. de Magalhães JP, Curado J, Church GM. 2009. Meta-analysis of age-related gene expression profiles identifies common signatures of aging. Bioinformatics. 25(7):875-881.

11. Zahn JM, Sonu R, Vogel H, Crane E, Mazan-Mamczarz K, Rabkin R, Davis RW, Becker KG, Owen AB, Kim SK. 2006. Transcriptional profiling of aging in human muscle reveals a common aging signature. PLoS Genet. 2(7):e115.

12. Glass D, Viñuela A, Davies MN, Ramasamy A, Parts L, Knowles D, Brown AA, Hedman AK, Small KS, Buil A et al. 2013. Gene expression changes with age in skin, adipose tissue, blood and brain. Genome Biol. 14(7):R75.

13. Voutetakis K, Chatziioannou A, Gonos ES, Trougakos IP. 2015. Comparative meta-analysis of transcriptomics data during cellular senescence and in vivo tissue ageing. Oxid Med Cell Longev. 2015:732914.

14. Yang J, Huang T, Petralia F, Long Q, Zhang B, Argmann C, Zhao Y, Mobbs CV, Schadt EE, Zhu J et al. 2015. Synchronized age-related gene expression changes across multiple tissues in human and the link to complex diseases. Sci Rep. 5:15145.

15. Zahn JM, Poosala S, Owen AB, Ingram DK, Lustig A, Carter A, Weeraratna AT, Taub DD, Gorospe M, Mazan-Mamczarz K et al. 2007. Agemap: A gene expression database for aging in mice. PLoS Genet. 3(11):e201.

16. Zhang B, Horvath S. 2005. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol. 4:Article17.

17. Hu Y, Pan J, Xin Y, Mi X, Wang J, Gao Q, Luo H. 2018. Gene expression analysis reveals novel gene signatures between young and old adults in human prefrontal cortex. Front Aging Neurosci. 10:259.

18. Li C, Mo D, Li M, Zheng Y, Li Q, Ou S, Zhang Z. 2019. Age-related but not longevity-related genes are found by weighted gene co-expression network analysis in the peripheral blood cells of humans. Genes Genet Syst. 93(6):221-228.

19. Li ML, Wu SH, Zhang JJ, Tian HY, Shao Y, Wang ZB, Irwin DM, Li JL, Hu XT, Wu DD. 2019. 547 transcriptomes from 44 brain areas reveal features of the aging brain in non-human primates. Genome Biol. 20(1):258.

20. Ferreira M, Francisco S, Soares AR, Nobre A, Pinheiro M, Reis A, Neto S, Rodrigues AJ, Sousa N, Moura G et al. 2021. Integration of segmented regression analysis with weighted gene correlation network analysis identifies genes whose expression is remodeled throughout physiological aging in mouse tissues. Aging (Albany NY). 13(14):18150-18190.

21. Consortium G. 2013. The genotype-tissue expression (gtex) project. Nat Genet. 45(6):580-585.

22. Pinheiro J, Bates D, DebRoy S, Sarkar D, Team RC. 2022. Nlme: Linear and nonlinear mixed effects models. R package version 3.1-155 ed.

23. Team RC. 2022. R: A language and environment for statistical computing.

24. Leek J, Johnson W, Parker H, Fertig E, Jaffe A, Zhang Y, Storey, JD, Torres L. 2021. Sva: Surrogate variable analysis. R package version 3.42.0 ed.

25. Langfelder P, Horvath S. 2008. Wgcna: An r package for weighted correlation network analysis. BMC Bioinformatics. 9(1):559.

26. Csardi G, Nepusz T. 2005. The igraph software package for complex network research. InterJournal. Complex Systems:1695.

27. Storey J. 2002. A direct approach to false discovery rates. Journal of the Royal Statistical Society Series B. 64:479-498.

28. Alexa A, Rahnenfuhrer J. 2022. *Topgo: Enrichment analysis for gene ontology*. R package version 2.50.0 ed.

29. Gehlenborg N. 2019. Upsetr: A more scalable alternative to venn and euler diagrams for visualizing intersecting sets. R package version 1.4.0 ed.

30. Langfelder P, Luo R, Oldham MC, Horvath S. 2011. Is my network module preserved and reproducible? PLoS Comput Biol. 7(1):e1001057.

31. Rodwell GE, Sonu R, Zahn JM, Lund J, Wilhelmy J, Wang L, Xiao W, Mindrinos M, Crane E, Segal E et al. 2004. A transcriptional profile of aging in the human kidney. PLoS Biol. 2(12):e427.

32. Lu T, Pan Y, Kao SY, Li C, Kohane I, Chan J, Yankner BA. 2004. Gene regulation and dna damage in the ageing human brain. Nature. 429(6994):883-891.

33. Palmer D, Fabris F, Doherty A, Freitas AA, de Magalhães JP. 2021. Ageing transcriptome meta-analysis reveals similarities and differences between key mammalian tissues. Aging (Albany NY). 13(3):3313-3341.

34. Rodrigues LP, Teixeira VR, Alencar-Silva T, Simonassi-Paiva B, Pereira RW, Pogue R, Carvalho JL. 2021. Hallmarks of aging and immunosenescence: Connecting the dots. Cytokine Growth Factor Rev. 59:9-21.

35. Batalha CMPF, Vercesi AE, Souza-Pinto NC. 2022. The many roles mitochondria play in mammalian aging. Antioxid Redox Signal. 36(13-15):824-843.

36. Jia K, Chen D, Riddle DL. 2004. The tor pathway interacts with the insulin signaling pathway to regulate c. Elegans larval development, metabolism and life span. Development. 131(16):3897-3906.

37. Vellai T, Takacs-Vellai K, Zhang Y, Kovacs AL, Orosz L, Müller F. 2003. Genetics: Influence of tor kinase on lifespan in c. Elegans. Nature. 426(6967):620.

38. Kapahi P, Zid BM, Harper T, Koslover D, Sapin V, Benzer S. 2004. Regulation of lifespan in drosophila by modulation of genes in the tor signaling pathway. Curr Biol. 14(10):885-890.

39. Kaeberlein M, Powers RW, Steffen KK, Westman EA, Hu D, Dang N, Kerr EO, Kirkland KT, Fields S, Kennedy BK. 2005. Regulation of yeast replicative life span by tor and sch9 in response to nutrients. Science. 310(5751):1193-1196.

40. Anisimov VN, Zabezhinski MA, Popovich IG, Piskunova TS, Semenchenko AV, Tyndyk ML, Yurova MN, Antoch MP, Blagosklonny MV. 2010. Rapamycin extends maximal lifespan in cancer-prone mice. Am J Pathol. 176(5):2092-2097.

41. Chen C, Liu Y, Zheng P. 2009. Mtor regulation and therapeutic rejuvenation of aging hematopoietic stem cells. Sci Signal. 2(98):ra75.

42. Harrison DE, Strong R, Sharp ZD, Nelson JF, Astle CM, Flurkey K, Nadon NL, Wilkinson JE, Frenkel K, Carter CS et al. 2009. Rapamycin fed late in life extends lifespan in genetically heterogeneous mice. Nature. 460(7253):392-395.

43. Lamming DW, Ye L, Katajisto P, Goncalves MD, Saitoh M, Stevens DM, Davis JG, Salmon AB, Richardson A, Ahima RS et al. 2012. Rapamycin-induced insulin resistance is mediated by mtorc2 loss and uncoupled from longevity. Science. 335(6076):1638-1643.

44. Miller RA, Harrison DE, Astle CM, Baur JA, Boyd AR, de Cabo R, Fernandez E, Flurkey K, Javors MA, Nelson JF et al. 2011. Rapamycin, but not resveratrol or simvastatin, extends life span of genetically heterogeneous mice. J Gerontol A Biol Sci Med Sci. 66(2):191-201.

45. Wilkinson JE, Burmeister L, Brooks SV, Chan CC, Friedline S, Harrison DE, Hejtmancik JF, Nadon N, Strong R, Wood LK et al. 2012. Rapamycin slows aging in mice. Aging Cell. 11(4):675-682.

46. Wu JJ, Liu J, Chen EB, Wang JJ, Cao L, Narayan N, Fergusson MM, Rovira, II, Allen M, Springer DA et al. 2013. Increased mammalian lifespan and a segmental and tissue-specific slowing of aging after genetic reduction of mtor expression. Cell Rep. 4(5):913-920.

47. Bjedov I, Toivonen JM, Kerr F, Slack C, Jacobson J, Foley A, Partridge L. 2010. Mechanisms of life span extension by rapamycin in the fruit fly drosophila melanogaster. Cell Metab. 11(1):35-46.

48. Powers RW, Kaeberlein M, Caldwell SD, Kennedy BK, Fields S. 2006. Extension of chronological life span in yeast by decreased tor pathway signaling. Genes Dev. 20(2):174-184.

49. Robida-Stubbs S, Glover-Cutter K, Lamming DW, Mizunuma M, Narasimhan SD, Neumann-Haefelin E, Sabatini DM, Blackwell TK. 2012. Tor signaling and rapamycin influence longevity by regulating skn-1/nrf and daf-16/foxo. Cell Metab. 15(5):713-724.

50. Bitto A, Ito TK, Pineda VV, LeTexier NJ, Huang HZ, Sutlief E, Tung H, Vizzini N, Chen B, Smith K et al. 2016. Transient rapamycin treatment can increase lifespan and healthspan in middle-aged mice. Elife. 5.

51. Shindyapina AV, Cho Y, Kaya A, Tyshkovskiy A, Castro JP, Deik A, Gordevicius J, Poganik JR, Clish CB, Horvath S et al. 2022. Rapamycin treatment during development extends life span and health span of male mice and. Sci Adv. 8(37):eabo5482.

52. Mazin P, Xiong J, Liu X, Yan Z, Zhang X, Li M, He L, Somel M, Yuan Y, Phoebe Chen YP et al. 2013. Widespread splicing changes in human brain development and aging. Mol Syst Biol. 9:633.

53. Adusumalli S, Ngian ZK, Lin WQ, Benoukraf T, Ong CT. 2019. Increased intron retention is a post-transcriptional signature associated with progressive aging and alzheimer's disease. Aging Cell. 18(3):e12928.

54. Tabrez SS, Sharma RD, Jain V, Siddiqui AA, Mukhopadhyay A. 2017. Differential alternative splicing coupled to nonsense-mediated decay of mrna ensures dietary restriction-induced longevity. Nat Commun. 8(1):306.

55. Heintz C, Doktor TK, Lanjuin A, Escoubas C, Zhang Y, Weir HJ, Dutta S, Silva-García CG, Bruun GH, Morantte I et al. 2017. Splicing factor 1 modulates dietary restriction and torc1 pathway longevity in c. Elegans. Nature. 541(7635):102-106.

56. Mariotti M, Kerepesi C, Oliveros W, Mele M, Gladyshev VN. 2022. Deterioration of the human transcriptome with age due to increasing intron retention and spurious splicing. bioRxiv.2022.2003.2014.484341.

57. Turi Z, Lacey M, Mistrik M, Moudry P. 2019. Impaired ribosome biogenesis: Mechanisms and relevance to cancer and aging. Aging (Albany NY). 11(8):2512-2540.

58. Shen L. 2022. *Geneoverlap: Test and visualize gene overlaps*. 1.34.0 ed.